



# THE BROOKINGS INSTITUTION

1775 Massachusetts Avenue, NW Washington, DC 20036-2103  
Tel: 202-797-6000 Fax: 202-797-6004  
[www.brookings.edu](http://www.brookings.edu)

## Innovations in Contextual Data for Crime Mapping

Andrew D. Reamer, Fellow  
Metropolitan Policy Program  
The Brookings Institution

Keynote Address  
Crime Mapping Research Conference  
Pittsburgh, Pennsylvania  
March 28, 2007

That I'm standing in front of you stems from a particular circumstance, the fact that Congress determines federal agency appropriations through a system that happens to lump the Commerce Department and the Justice Department in the same appropriations bill. The story goes like this:

At Brookings, my job is to see that the federal government provides the local area statistics that governments at all levels and the private sector need for intelligent decision-making and investment. The federal statistical system is idiosyncratic and decentralized, with responsibilities divided among a fair number of federal agencies. For the purposes of the Brookings Metropolitan Policy Program, the foundational statistical agency is the U.S. Census Bureau, which provides data on population size and characteristics down to the neighborhood level, and the Census Bureau's foundational program is the Decennial Census, which counts the people in every household in America on the first of April in the year ending in zero. The Decennial Census is critical—for Congressional apportionment, determining the flow of federal funds, and aiding government and business decision-making.

The Decennial Census, you might imagine, is an enormous undertaking. For the 2010 Census, the cost of planning, implementing, analyzing, and disseminating will be north of \$11 billion. While the Census Bureau has many statistical responsibilities, the Decennial Census is by far the biggest, so the agency budget waxes and wanes with the chronological distance from the year ending in zero.

The amount of federal funds available for domestic policy initiatives has been tight for some time. So when the Administration asks for a multi-million dollar increase in the Census Bureau budget every year we move towards the Decennial Census, some Congressional mouths water. Every June, when the Commerce, Justice, and State budget hits the House floor, various members offer amendments to shift funds from "those bureaucrats" in the Census Bureau to programs that aid "the hard-working men and women of our state and local law enforcement agencies." Last year, Representatives put forward seven amendments to take money from

Census, the 2010 Census in particular. Anticipating this in light of similar events in 2004 and 2005, I wrote an editorial saying that an adequate Census Bureau budget was necessary to fight crime—that census data are instrumental in crime analysis and, moreover, the appropriate distribution of federal assistance to state and local criminal justice efforts actually depends on having good census data. Ron Wilson saw that article, learned about my program at Brookings, and here I am.

Ordinarily, the closest I get to criminal justice is that my social work professor brother is on the Rhode Island Parole Board and writes academic books about the types and causes of crime. But with a background in urban planning and economic development, I do know something about the sources and uses of small area statistics. Ron thought it would be useful for me to give you an overview of the breadth of data resources beyond the traditional ones used in crime mapping so that you can be even more relevant and effective in your work, whether you're a practitioner or a researcher, and to suggest steps that you can take individually, organizationally, and as a field to take advantage of these data sources.

The development and widespread dissemination of innovative, useful crime mapping techniques has been made possible by relatively recent advances in information technology hardware and software—particularly our ability to electronically manipulate and analyze large amounts of data and prepare visual representations of the results. Experts in a wide variety of fields, from health care to transportation to education to economic development to criminal justice, eagerly have made use of these IT advances. But, understandably, this has been in a somewhat siloed fashion, with each policy realm unto its own. With so much happening so quickly, efforts to talk across silos happen too rarely. Consider this talk one effort at building bridges across realms.

As I understand it, the particular purposes of crime mapping are several:

- to more efficiently allocate scarce crime fighting resources,
- to identify more effective interventions for preventing and halting particular types of crime, and
- to influence city, regional, and state non-criminal justice policy matters (e.g., housing, transportation, education, community development) that affect criminal activity.

Moreover, I understand that crime mapping aids these processes through several roles:

- describing patterns and correlations regarding who, where, when, and how,
- explaining these patterns,
- predicting events and prescribing interventions on the basis of these explanations, and
- evaluating the results of these and other crime-fighting activities.

The foundation of the crime mapping enterprise is the first part, describing patterns and correlations. And the field recognizes, I see from looking at the MAPS site, that exploring the correlations between crime data and socioeconomic data can do much to improve the value of the analysis. This is true whether you are in a local police department or an academic research institution.

As I'll show you, the world of small area statistics is rapidly changing for the better, in a number of ways.

- First, we're the beneficiary of a new and transformative federal survey, the American Community Survey, which will provide annually updated, highly detailed population characteristics down to the block group level.
- Second, we now have tools and procedures that provide us with on-line access to massive local, state, and nationwide administrative datasets valuable for crime mapping.
- Third, in the last few years, statistical techniques have advanced to the point where we are now able to create small area synthetic data, of two types.
  - The first is synthetic microdata—data that mimic real, but confidential, microdata and produce the same statistical results in terms of mean, median, variance and so forth.
  - The second is synthetic aggregate data for small areas, developed through marrying a small national dataset on one topic, say household transportation patterns, with one that provides demographic data at the neighborhood level, like the long form of the census.
- Fourth, we are well along in developing web sites and tools that provide access to and integration of datasets from multiple sources. While most of these house the various datasets on one central server, a handful are distributed data systems, pulling data on the fly directly from each of the individual data sources.

I'm about to take you on a reconnaissance of specific data efforts that reflect these various trends. I'm going to frame this reconnaissance around four types of actions that I encourage you to take after this conference: explore, link, collaborate, and advocate. That is:

- ***explore*** a wide variety of new, interesting datasets and tools;
- ***link***, or make use of, the ones of particular value to you for describing, explaining, prescribing, and evaluating;
- for those who have the time and resources, ***collaborate*** with those in different policy realms, in adjacent geographic areas, and in different levels of government, so you might obtain develop more valuable datasets; and
- ***advocate***, ask federal, state, and local government decision-makers to support the datasets and data programs important to your work.

I'll begin by discussing a variety of federal datasets and tools, and move on to local, state, and other. You can relax a bit on the notetaking, as I'll be working with Ron to create a web page on the MAPS site that describes and links to these various items.

## Federal

Traditionally, we essentially relied on one federal dataset for small areas—the decennial census. For the last 60 years, the decennial census had two parts, the short form that five of six household filled out, and the long form, which the other one-sixth got. We can throw out the past now. Due to the various trends I mentioned, we are in the midst of a significant restructuring and expansion of federal efforts that produce small area socioeconomic data.

The American Community Survey – In my outsider’s view, perhaps the single most important innovation in federal statistics for the purposes of crime mapping is the transformation of the long form survey carried out once a decade into the continuously carried out, annually updated American Community Survey (ACS). The ACS survey instrument is almost exactly the same as that for the long form, collecting data on educational attainment, citizenship, ancestry, language, residential mobility, employment status, industry and occupation, income, journey to work, and housing characteristics. 250,000 households across the nation are surveyed each month, 3 million a year, and the data for each calendar year are released just eight months later, at the end of August. Now the Decennial census long form went to 17 million households, so to build sufficient sample size for small areas, we’ll need several years’ worth of data. Areas with population of 65,000 or more will get single-year estimates, the first of which came out in 2006 for 2005; areas of 20,000 to 65,000 will get annually updated three-year estimates, the first being 2005-2007, released in 2008; areas with less than 20,000, down to the census tract and block group levels, will get annually updated five-year estimates, the first being 2005-2009, released in 2010.

If you know the Decennial Census long form, you will see a few new wrinkles in the ACS. The ACS asks respondents where they lived a year ago, rather than five years ago; going back one year is probably more valuable for crime analysis. In 2008, a question on health insurance will be added. In 2006, the ACS began to survey group quarters, such as halfway houses and prisons.

As I said, I think that for crime mappers, the ACS is probably the most valuable statistical program innovation to come along in some time. And as we’ll see, there’re others that have great value as well.

Before discussing those, I do need to mention the 2010 Census, which will be short form only, giving us population size and the basic characteristics of age, gender, race, and Hispanicity down to the census tract and block group levels. The 2010 Census data will be useful for crime analysis, of course, but they take on even more importance because the population estimates for the ACS and other demographic surveys are tied to the decennial count.

Recognize that an inaccurate population count will have consequences for your work. The Census Bureau’s good at decennial census data collection, it’s been doing this job for 100 years or so, but it’s not perfect and it needs state and local government help, in two ways. One is to review and correct the census address list—if an address is missing, it doesn’t get a survey. Local governments can draw on a variety of resources, 911 records for instance, to improve the address list. The second way local governments can help is what’s being called “getting out the count,” encouraging various hard-to-reach populations to fill out the surveys. If you want effective crime

analysis, it's in your interest to encourage your state and local governments to help out the Census Bureau in these ways.

Administrative Datasets. The second realm of innovation in federal statistics, emerging over the last eight years or so, is web access to large public use databases that let you drill down to the census tract level or lower. I'll mention three examples from the realm of housing to give you a sense of the possibilities. It seems to me that data on housing and mortgage characteristics have some usefulness in crime analysis, particularly in explanation, prediction, and evaluation.

The HUD Office of Policy Development and Research (PD&R) provides several housing-related databases. One is the Government Sponsored Enterprises (GSE) public use database. This provides data on mortgages purchased by Freddie Mac and Fannie Mae for single and multi-family residences. For single family housing, data are available annually by census tract for the number of borrowers, their race, gender, age, their income levels, the loan characteristics, and the number of first-time home buyers.

PD&R's Low Income Housing Tax Credit (LIHTC) database provides information on nearly 25,500 projects and over 1,415,000 housing units placed in service since 1987. The LIHTC is used to support the acquisition, rehabilitation, or new construction of rental housing targeted to lower-income households.

The Federal Financial Institutions Examination Council (FFIEC), which is made up of the various federal bank regulators, produces an annual on-line database of home mortgages granted and denied for each census tract by type of loan, race, and income. These data are collected through the workings of the Home Mortgage Disclosure Act (HMDA).

Synthetic Datasets. Now we move a bit closer to the cutting edge, synthetic datasets. I'm going to discuss two examples concerning local travel. I mentioned we now have the ability to create synthetic microdata—data that mimic real, but confidential, microdata and produce the same statistical results in terms of mean, median, variance and so forth.

The Census Bureau's Local Employment Dynamics (LED) program links three massive federally-funded datasets—business establishment data (how many people work where, in what industry, and how much they get paid) with employee wage records (each person's job and pay history) and a Social Security file. (The first two files comes from the unemployment insurance system.) Through linking these very large administrative datasets, LED has created an On The Map tool that visually shows where people live in relationship to where they work. To prevent disclosure of confidentiality, LED is using a synthetic microdata set that mimics the actual data. This allows you to draw your own boundaries with a mouse, as small an area as you like, and produces tables based on those data. A variety of analyses is possible using these data, for example, daytime populations for areas of interest as well as journey-to-work transportation flows. On The Map is now available in 17 states; the number will go to 44 states by September 2007.

At a somewhat slower pace, LED is making public use synthetic microdata sets available. The data package consists of origin-destination data; residence area characteristics data, such as

gender and age; workplace area characteristics data, such as industry and pay; and block-group level indicators on items such as hires and fires (job market churning). Currently, data are available for Oregon, Texas, Washington, Virginia, Colorado, and Florida, with more to come.

I think I can say that the LED On The Map PUMS has the potential to be transformative for a variety of fields, including crime mapping. There's a great dissertation here waiting for someone.

The National Household Transportation Survey (NHTS) illustrates a second type of synthetic dataset potentially useful for crime mapping. NHTS is carried out by the U.S. Department of Transportation every six or seven years and describes local and long distance travel patterns by who, how, when, how often, how long, and why. In 2001, the NHTS sample for the nation was 25,000; state purchased add-ons brought the number to 65,000, still far too small to do anything meaningful at a local level. But some enterprising researchers at the Federal Highway Administration and Oak Ridge National Laboratories created what they dubbed a transferability model. They married the 1995 NHTS sample with 1990 Census long form data to create synthetic estimates of local travel patterns for homogeneous census tract clusters across the nation. I believe the 2001 transferability model is to be released by June. The next NHTS, funding willing, is scheduled to be carried out in 2008.

Federal Data Access Tools and Resources. IT advances have enabled not only greater availability but improved accessibility as well, in two ways. First, there are several one-stop web sites that allow you to download publicly available data from multiple sources, federal and otherwise.

The first is the Census Bureau's Data Ferrett, which gives you access to a wide variety of federal summary data and public use microdata for the geographies you specify, down to the census tract level. Data Ferrett allows you to obtain an integrated data set that pulls all these sources together into one output, and you can upload your own data so it's in the mix as well.

The Federal Geographic Data Committee's Geospatial One-Stop provides direct access to a variety of federal, state, local and other data sets, maps, and other resources.

On the private, nonprofit side of the world, the Fannie Mae Foundation's DataPlace.Org has pulled together a variety of federal demographic and housing data, and allows you to create maps, charts, and graphs for standard and customized geographic areas down to the census tract level. For instance, you can group census tracts together with boundaries that make sense to you. As with Data Ferrett, you will soon be able to upload your own data as well.

The second type of access tool lets people work with actual confidential microdata. For about a decade, the Census Bureau has operated a series of Research Data Centers (RDCs) across the country that let researchers with the proper clearance to work with confidential microdata on the Census Bureau's computers. This is good, but the process is cumbersome, and you have to be physically at one of the sites, which are in Massachusetts, New York, California, DC, Michigan, Illinois, and North Carolina. And if you're not close to one of them, too bad.

Happily, the Census Bureau is creating a Remote Access System which will let researchers from anywhere work with ACS microdata from outside the firewall. You can specify data runs and get

the results without touching the actual data. The Remote Access System is in beta phase, and should be quite a boon to those wanting to work with the ACS. Census will be looking for beta testers, so this is an opportunity to collaborate with the Bureau.

Before moving on from federal to other datasets, I want to make a pitch for advocacy, the fourth of my four suggested actions. I've covered datasets provided through the Census Bureau, HUD, and the Departments of Labor and Transportation. The statistical budget for every one of these agencies has been cut, drastically in the cases of HUD and Transportation, and will be targeted for cuts in the future. Congress and the Administration both have difficulty recognizing that relatively small investments in statistical programs, a few million here and there, have an enormous return to the nation. So I encourage you to write your representatives in Congress and let them know about the federal data programs that are important to your work.

I'll wrap up by briefly touching on datasets from national research institutes, state governments, and community information systems.

### Datasets from National Research Institutes

The Urban Institute's Reentry Mapping Network (RMN) maps the reentry of prisoners to communities in 15 metropolitan areas across the nation. This dataset allows researchers to examine the correlation between prisoner reentry into society and crime.

### State Datasets

State government datasets are an underappreciated resource for the purposes of public policy. While data availability and accessibility will vary from state to state, it's fair to say that each state has administrative and survey datasets that can be useful to your purposes.

You will need to go department by department – housing, workforce, employment, health, environment, transportation, education—to see what's available. At the same time, recognize that each state participates in a number of federal-state cooperative data systems—in employment, health, transportation, education, and, I imagine, criminal justice, so a number of relatively standardized datasets will be available in each state.

### Local Statistics

Over the last decade, IT advances have enabled the bottom-up development of community information systems across the nation that provide integrated access to a variety of administrative and survey datasets from local, state, and federal sources. Examples of local administrative data sources include tax records, housing enforcement, 911 calls, emergency medical services, sheriff's department service of notices, reports of criminal activity, public schools, day care licenses, child welfare records, parking permits, and liquor licenses. Many of these community information systems provide on-line tools for mapping and download. They have significant potential value for crime mapping, it seems to me.

The sponsors of these systems are varied, from local universities to non-profit community foundations to local governments. While these resources were enabled by technical advances, they are possible only through institutional trust, as all these various local agencies are handing over their data and they want some assurance the data won't be misused. Here's an overview of some developments nationally:

- Since the late 1990s, the Urban Institute has managed the National Neighborhood Indicators Partnership (NNIP), a trade association of organizations that run community information systems in 29 cities. NNIP promotes problem-sharing and solving among its members, and expects to add four new cities a year.
- The Boston area's metropolitan council has created an on-line DataCommon that brings together data on a multiplicity of topics for 101 cities and towns, an area far larger than most community information systems. Imagine the work it takes to standardize the data from 101 cities and towns.
- Most community information systems put their data on a single server. The next generation is a distributed community information system, for which the data reside at the original sources, not a central server. Someone accessing the data from the web interface would see data tables pulled on the fly from a variety of sources in real time. In the Chicago area, the Illinois Data Exchange Affiliates (IDEA) is trying to bring this concept into being.
- And here in Pittsburgh, the idea of a distributed system is being taken in an additional direction. Humanservices.net is a distributed Information Commons that allows users to map the offerings of 3,000 social service institutions and 10,000 social service programs in Allegheny County. The information available is updated by each institution directly.

In the realm of community information systems, the opportunities to explore and link are clear. And so are the opportunities to collaborate with other agencies and adjacent communities, and to advocate for the creation of such a system in your area if one doesn't exist and for the expansion of any existing system.

In addition to information systems, many communities are building on-line neighborhood indicator systems, which pull, manipulate and analyze data from a variety of sources with a focused purpose, to evaluate neighborhood conditions on topics from teen pregnancy to environmental pollution to the quality of local parks. Crime mappers may find neighborhood indicators particularly valuable as you seek to explain patterns and evaluate results.

In conclusion—there's a lot going on out there at the national, state, and local levels. There's an important new survey, the ACS, big administrative datasets, small area synthetic data, and tools that greatly expand our ability to access and integrate. I strongly encourage you to explore this great new world, link up with those efforts that improve your crime mapping, collaborate with new partners to build new tools and relationships, and advocate for adequate funding at the federal level and robust community information systems at the local level.

I hope you found this information of value. We'll place links to these various efforts on the MAPS web site. Thank you.